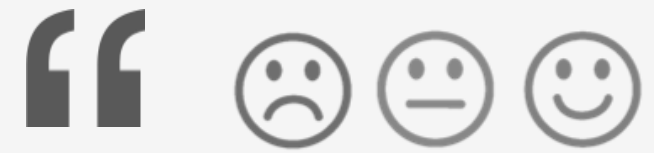


NLP 감정분석 기반
마케팅 시장 분석

온라인 커뮤니티 NLP 감성 분석 : 신조어 추출 및 분석

빅데이터미네이터



기존 계획: 커뮤니티별 주요 이슈 키워드 감성 분석 시스템”

커뮤니티 별 일일 키워드 추출 및 감성분석

→ 각 커뮤니티에서 키워드 추출

→ “신조어 분석”

→ 키워드별 연관 단어, 나타난 게시물 수, 사용된 문장 예시, 전체 감성 분석 수치 및 시각화 제공하는 프로그램 구현

주제 수정



온라인 커뮤니티 NLP 감성 분석 활용을 위한, 커뮤니티 신조어 추출 및 분석

커뮤니티 게시글의 키워드 추출 및 감성분석의 정확도에 큰 영향을 미치는 "신조어 추출 및 분석" 단계를 전체 주제로 확정.

• 기존 방식의 문제

- Voting 방식으로 일일이 신조어 및 그 극성 수치를 추가해야 함 (수동적)
- 신조어는 범용 감성 사전에 없는 경우가 많아, 감성 분석 결과가 중립으로 분류됨
- 범용 감성 사전에 존재하지 않는 신조어의 경우 전처리 단계에서 삭제됨



기대 효과

신조어를 자동으로 추출하고 해당 단어의 극성 및 점수를 제공한다면, 기존 온라인 데이터 감성 분석에 있어서 중립으로 추출되던 신조어들에 대한 극성 데이터도 포함되어 더 정확도 높은 결과를 기대할 수 있다.

조건 수정

온라인 커뮤니티 NLP 감성 분석 활용을 위한, 커뮤니티 신조어 추출 및 분석

- 기존의 최종 결과물인 카카오톡 챗봇 계획 삭제
- 팀원 부족 및 주제 수정으로, 원활한 테스트를 위해 커뮤니티를 2곳으로 확정
 - 네이트판 (방송 관련 신조어, 10대 게시판 등 신조어가 많이 나오는 게시판 다수 존재)
 - FM 코리아 (기존에 선정한 타 커뮤니티에 비해 유머 글이 많이 올라오는 게시판 존재)

nate 

착한펍코
에펍코리아

커뮤니티 신조어 추출 및 분석 계획

- 3-5일 등 기간을 정한 뒤 데이터 크롤링 진행, 해당 데이터에서 커뮤니티 신조어 후보 추출
- 후보 신조어 추출 후, 딥러닝 적용하여 실제 많이 사용되는 지 등 중요도로 분류하여 최종 신조어 채택 (차주 카카오 khaiii 등 딥러닝 모델을 활용하여 모델 구축 및 테스트 진행 예정)
- Word2Vec 등 기존 방식 보다 연관 단어 추출의 정확도가 높은 Fasttext 라이브러리를 이용하여 최종 신조어에 관한 주변 단어를 명사 단위로 추출하고 그 단어들의 극성을 활용하여 최종 신조어 극성 값 추출
 - 연관 단어 분석 시 KNU 감성 사전 (범용 사전) 이용
 - 추출된 신조어 극성 값은 최종적으로 타 범용 사전 혹은 개인이 구축한 특화 사전 등에 추가하여 감성 분석의 정확도를 높이는 것에 활용 가능하게 하는 것이 목표

차주 계획

1. 신조어 추출에 사용할 게시판 재선정
2. 사용할 데이터 크롤링
3. 신조어 추출 모델 구현 및 Test



감사합니다 :)
빅데이터미네이터